Universidad Popular del Cesar

Minería de datos y Patrones

Version 2025-I

Presentación

Dr. José Ramón Iglesias
DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingenieria Electronica
Universidad Popular del Cesar

Introducción

Definición

La minería de datos tiene como objetivo la extracción de conocimiento a partir de grandes cantidades de datos, mediante métodos automáticos o semiautomáticos.

Propone usar un conjunto de algoritmos [...] para construir modelos a partir de los datos, es decir, encontrar estructuras interesantes o patrones según criterios predefinidos, y extraer un máximo de conocimientos de ellos.[1]

[1] https://fr.wikipedia.org/wiki/Exploration_de_donn´ees

Datos

Diferentes tipos de datos:

- Datos estructurados:
  - Datos sociales: Edad, Salario, Color de piel, Lugar de residencia
  - Datos métricos: *Likes* de una publicación, Tiempo pasado en una página, Número de amigos en común
- Datos no estructurados:
  - Texto: Frase, Párrafo, Documento
  - Sonido: Canción, Discurso
  - Imagen: Foto, Vídeo

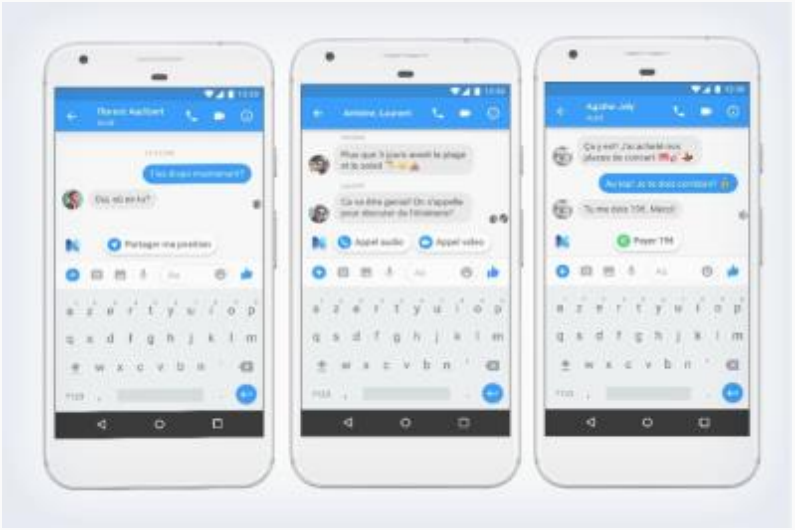
Diferentes tipos de Minería:

- Exploración de datos: Detectar valores simples, sesgos
- Tarea de clasificación/regresión: Alimentarse de datos para caracterizar nuevos datos por clase o con un valor, de manera supervisada
- Tarea de agrupamiento: Caracterizar datos por clase de manera no supervisada
- Reducción de dimensiones: Desarrollar estructuras comunes para representaciones comprimidas de datos

Outline : Aplicaciones

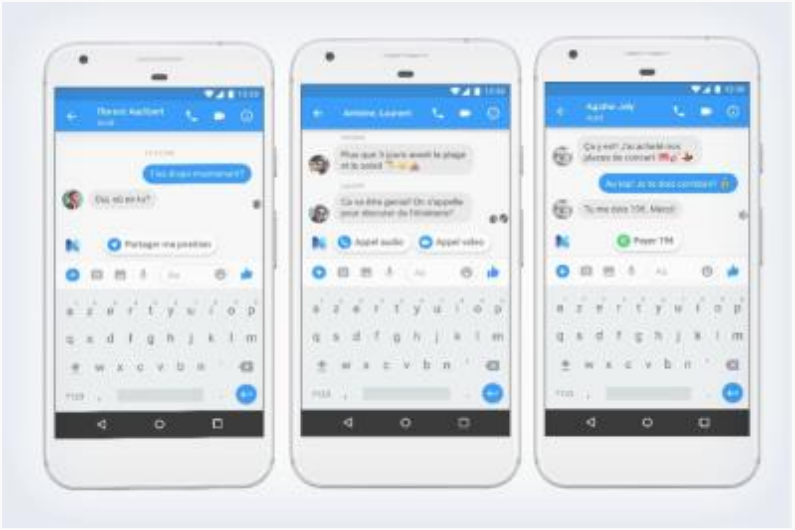
Aplicaciones (I/II)

- Detección de eventos en un texto



Aplicaciones (I/II)

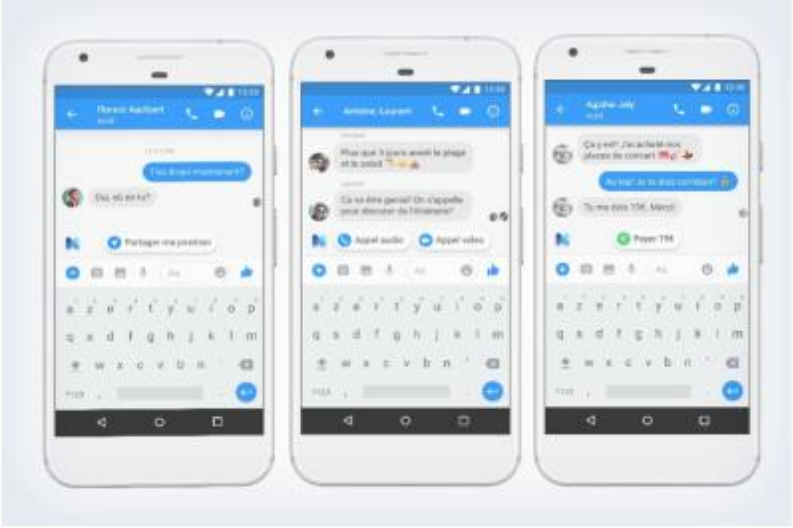
- Detección de eventos en un texto



- Procesamiento automático de opiniones de usuarios

Aplicaciones (I/II)

- Detección de eventos en un texto

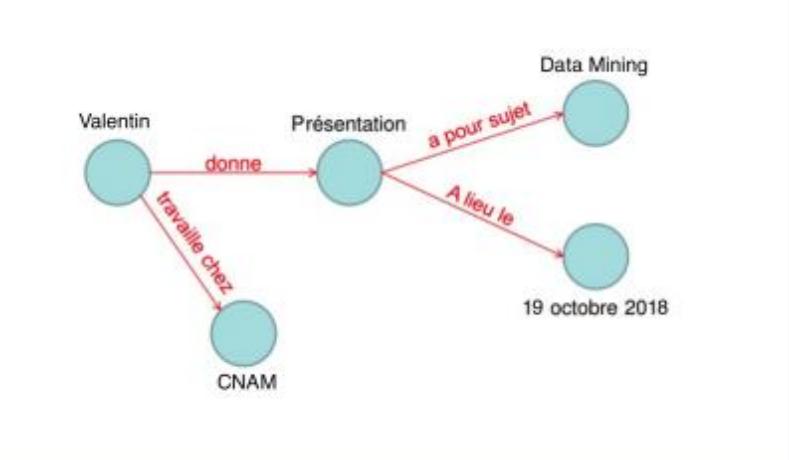


- Procesamiento automático de opiniones de usuarios

- Propuesta de recomendaciones a un usuario

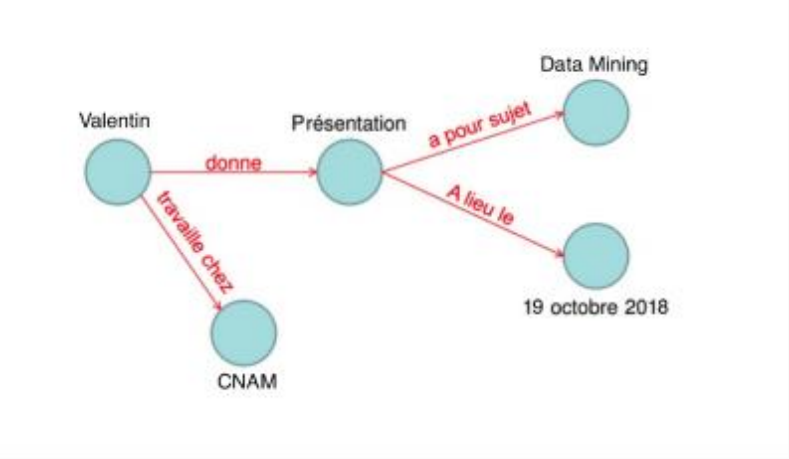
Aplicaciones (II/II)

- Detección de relaciones entre entidades en un texto

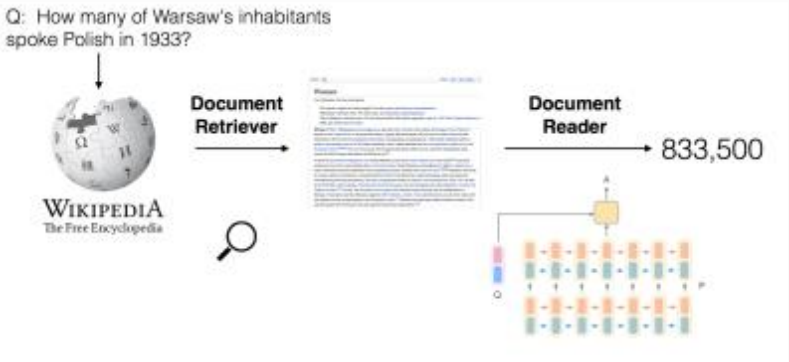


Aplicaciones (II/II)

- Detección de relaciones entre entidades en un texto

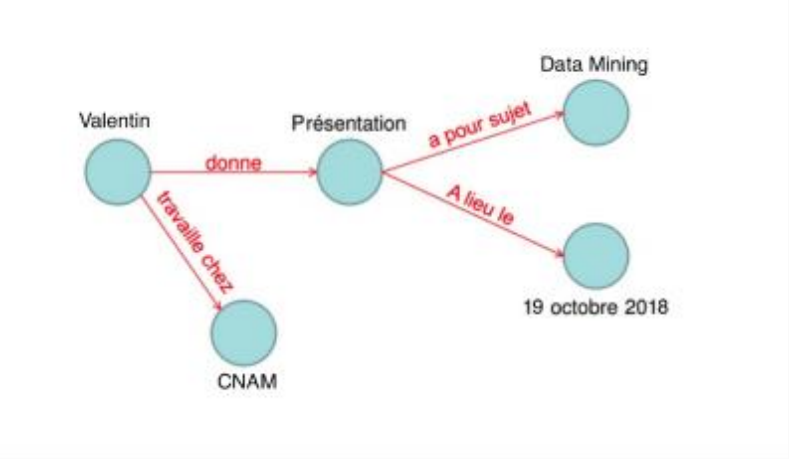


- Respuesta a una pregunta

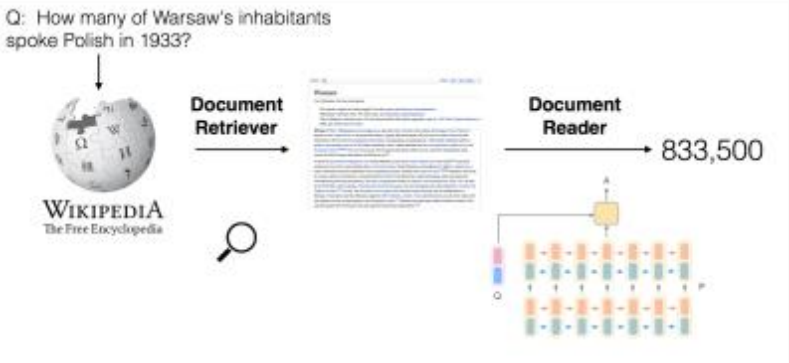


Aplicaciones (II/II)

- Detección de relaciones entre entidades en un texto



- Respuesta a una pregunta



- Módulo de IE para un agente conversacional

Outline : Significacion de las termas

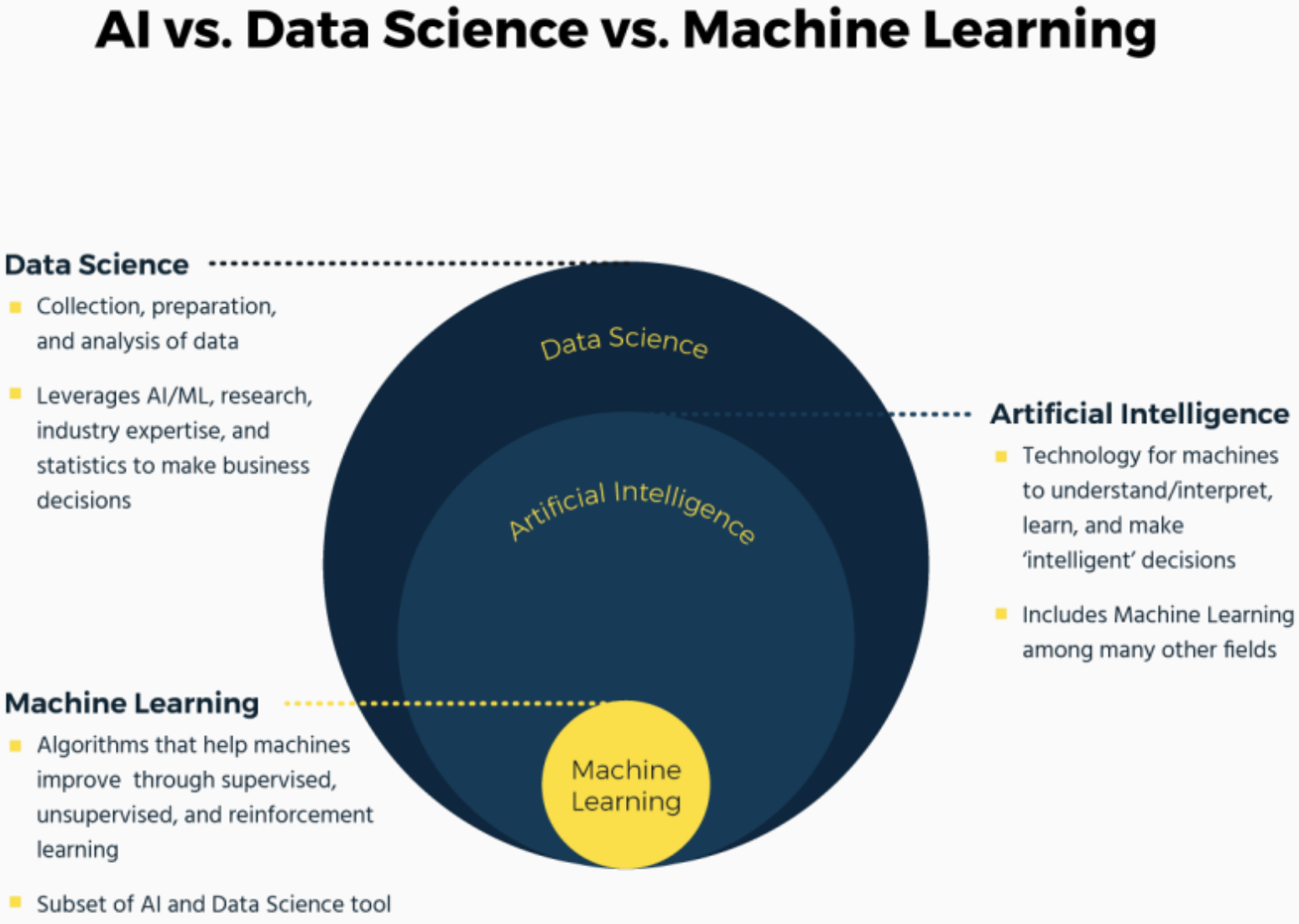


Figure 1: Diferencias entre campos

Data Science

En resumen

Data Science se centra en el análisis de datos para extraer conocimiento, Machine Learning utiliza algoritmos para hacer predicciones y tomar decisiones basadas en datos, y Artificial Intelligence se refiere al desarrollo de sistemas que pueden realizar tareas inteligentes de manera autónoma.

Definición (sobre) simplista:

- Data mining genera entendimiento.
- Machine learning genera predicciones.
- Artificial intelligence genera acciones.

Ejemplo en plataforma de musica

Data Scientist

Recopila y analiza datos de usuarios de plataformas de música para identificar patrones y preferencias musicales.

Machine Learner

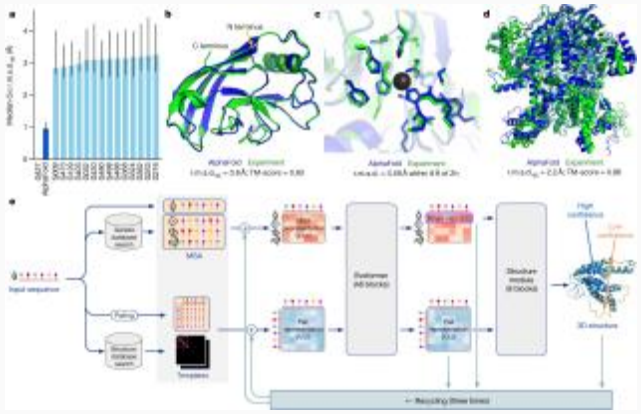
Desarrolla y optimiza un modelo de recomendación de música utilizando algoritmos de aprendizaje automático para predecir las preferencias de los usuarios.

Artificial Intelligence

Implementa un agente social que puede interagir con el usor, para mejorar la personalización de las recomendaciones musicales y proporcionar una experiencia más precisa y contextualizada.

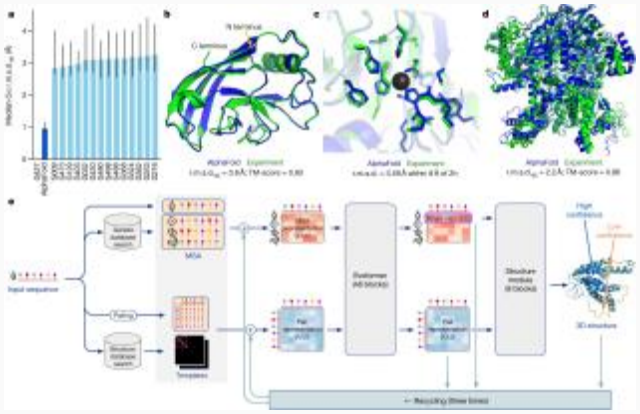
Significacion del trabajo: Porque hacer eso?

- Avance científico



Significacion del trabajo: Porque hacer eso?

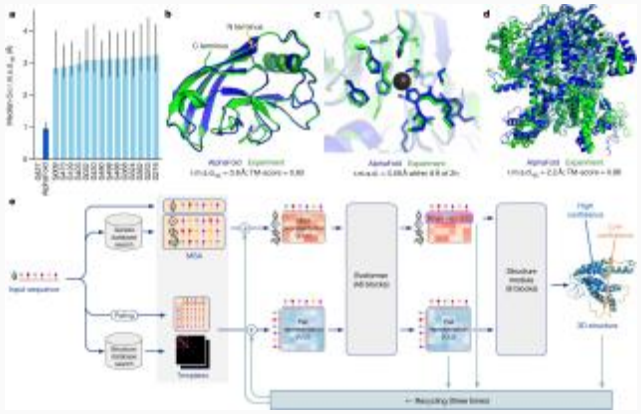
- Avance científico



- Prevención y gestión de desastres naturales

Significacion del trabajo: Porque hacer eso?

- Avance científico

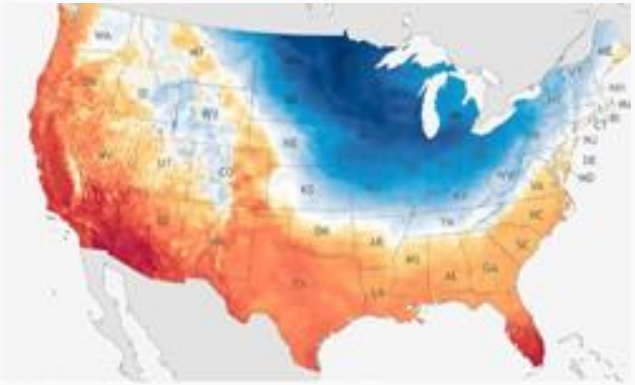


- Prevención y gestión de desastres naturales

- Impacto en la salud pública

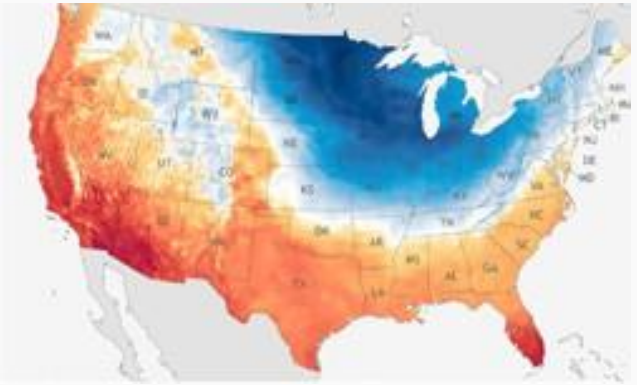
Significacion del trabajo: Porque hacer eso?

- Sostenibilidad ambiental

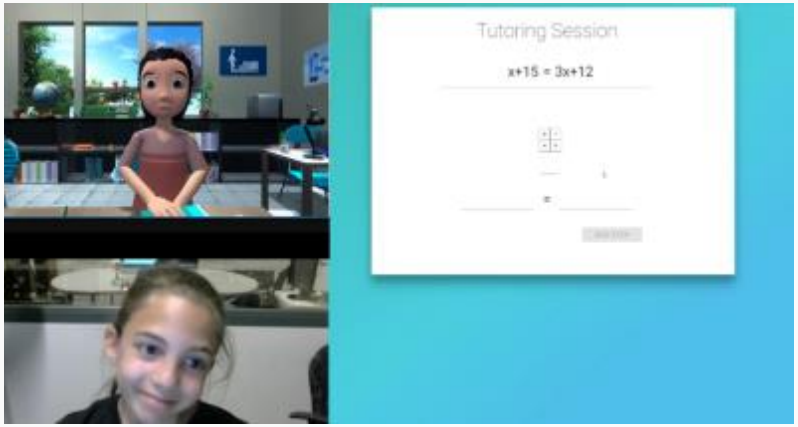


Significacion del trabajo: Porque hacer eso?

- Sostenibilidad ambiental

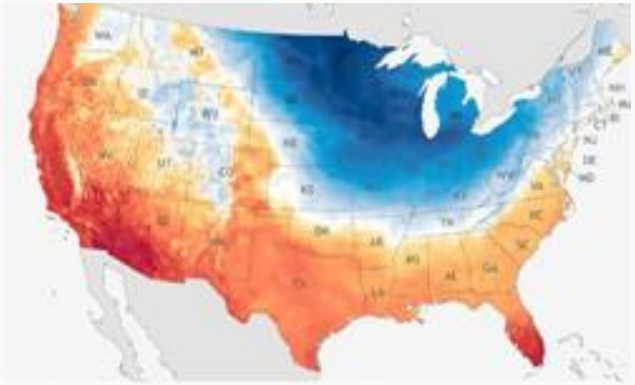


- Impulso a la educación y la investigación

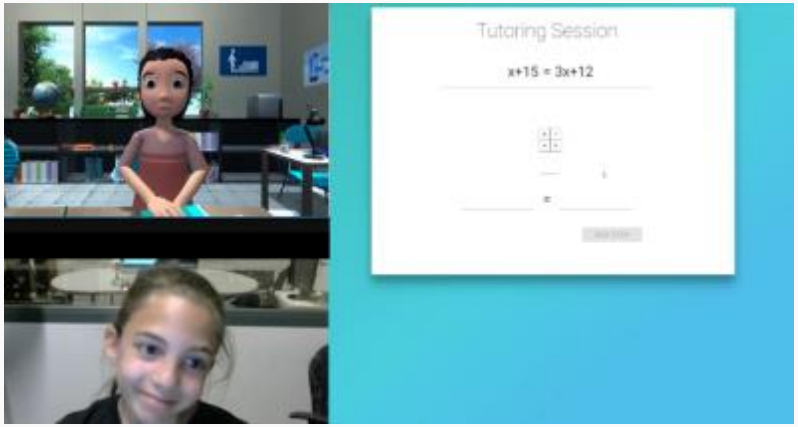


Significacion del trabajo: Porque hacer eso?

- Sostenibilidad ambiental



- Impulso a la educación y la investigación



- Democracia participativa

Outline : Prerrequisitos

Aplicaciones

Significacion de las termas

Prerrequisitos

TP Exploración de Datos: MovieLens

Overview

El programa

Partes teóricas

Bases de estadística, álgebra lineal: Presentación general del aprendizaje estadístico, Bases matemáticas de los diferentes modelos, enfoque intuitivo

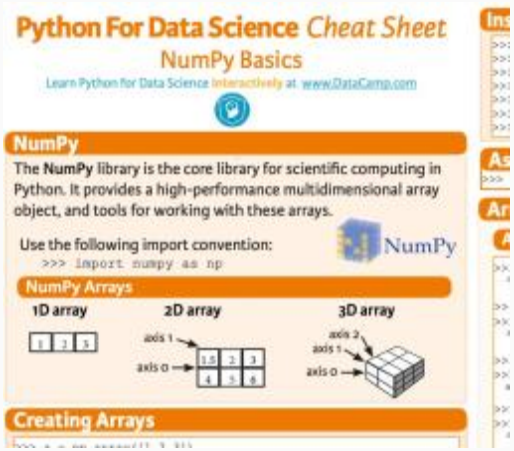
Partes prácticas

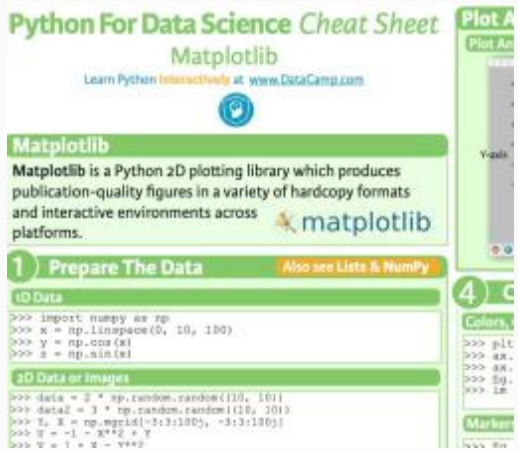
Bases de Python: Uso de herramientas de manipulación de datos, Uso de una biblioteca de DL, Uso de una biblioteca de ML, Análisis de sentimientos, Ranking sobre preferencias de vino, ...[2]

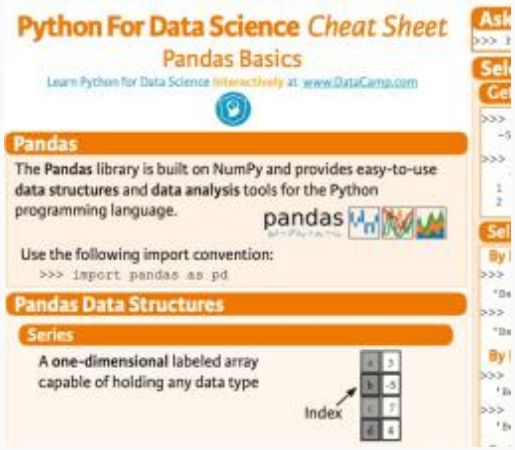
[2]Non contractual por este curso

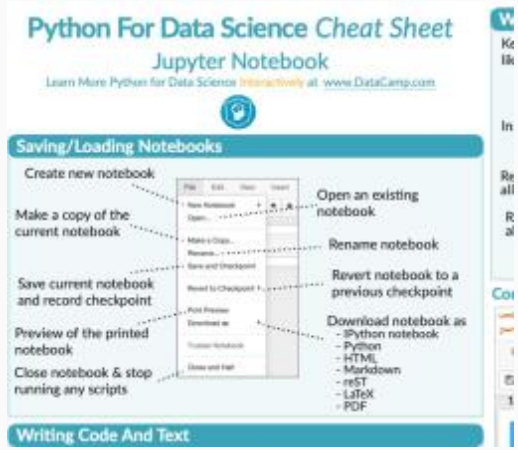
Material

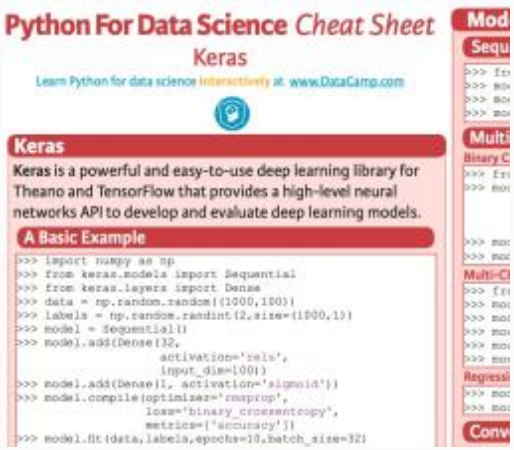
- Computadora
- Jupyter Notebook y Anaconda:
  https://www.anaconda.com/download/
- Los notebooks y las cheatsheets disponibles online:











Outline : TP Exploración de Datos: MovieLens

Exploración de Datos: MovieLens

Estudio simple de un conjunto de datos de críticas de películas

- 3 millones de notas
- Descriptores sociales: edad, sexo, ...
- Primer enfoque básico de minería de datos

movielens

Introducción a pandas

- Biblioteca de Python para manipular bases de datos: https://pandas.pydata.org/
- Permite realizar operaciones y visualizaciones
- Fácil de usar

Outline : Overview

Aplicaciones

Significacion de las termas

Prerrequisitos

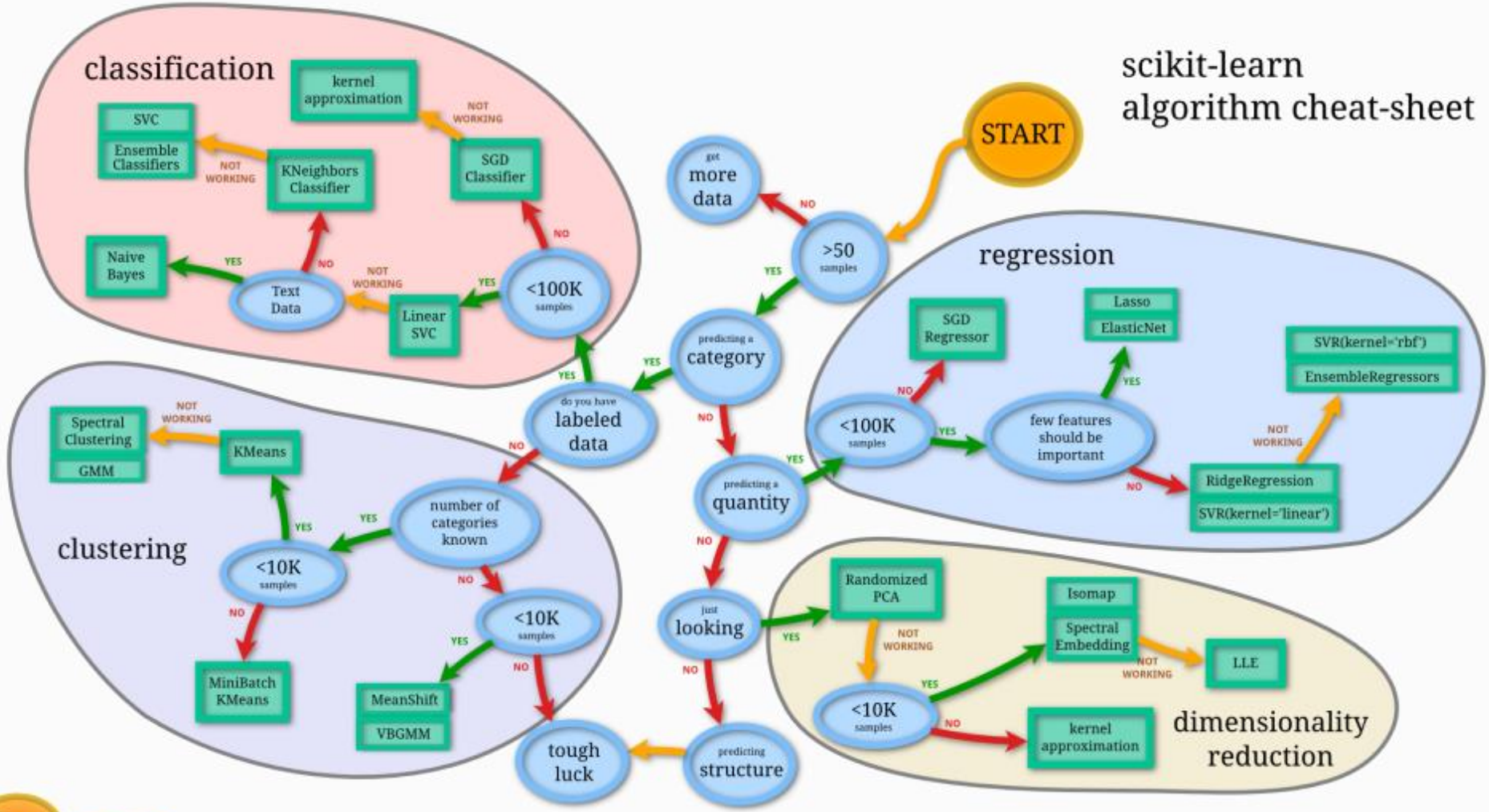
TP Exploración de Datos: MovieLens

Overview

Los diferentes métodos

- Clasificacion: predicir una clase determinadaa
- Regresion: predicir una valor
- Clustering: grupear elementos en cluster (clases non determinadas)
- Reduccion de dimension: diminuir el espacio de representacion de los datos
- Deteccion de anomalia

Los diferentes métodos



Back

Ejemplos: Classificacion

- Reconocimiento de emociones en el habla:
- Clasificación de especies de animales a partir de imágenes:
- Detección de objetos en imágenes médicas:

Ejemplos: Classificacion

- Reconocimiento de emociones en el habla: **la persona esta enojada o feliz**
- Clasificación de especies de animales a partir de imágenes:

- Detección de objetos en imágenes médicas:

Ejemplos: Classificacion

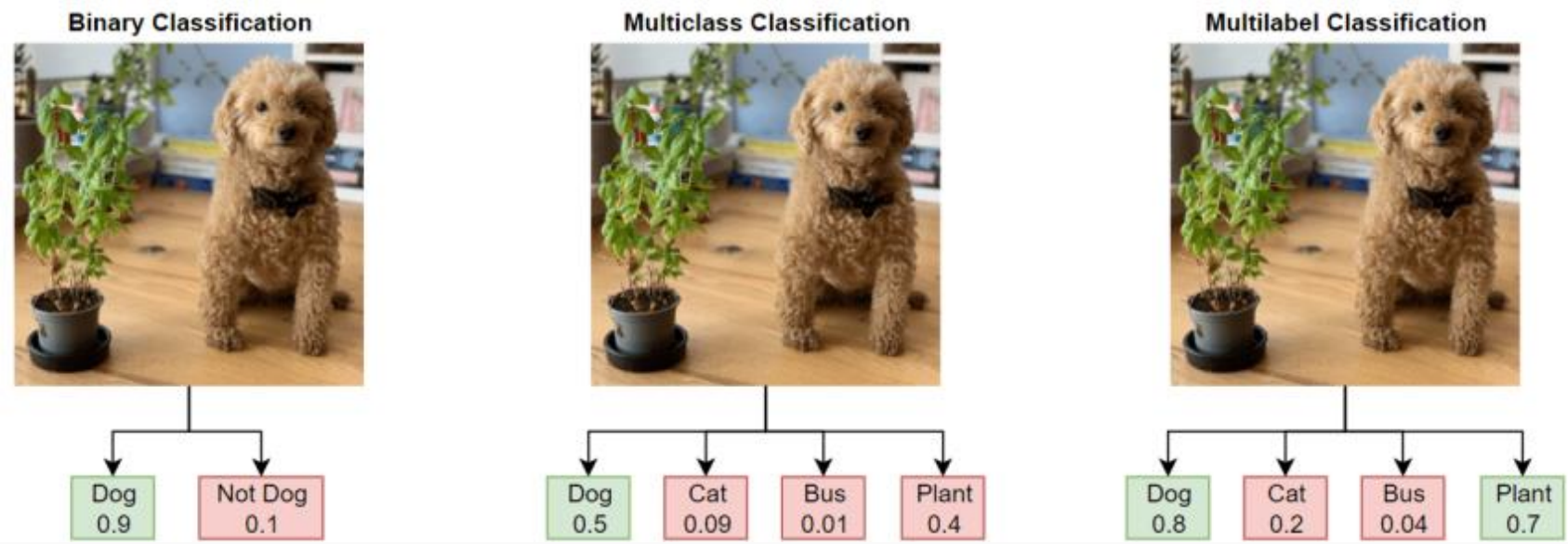
- Reconocimiento de emociones en el habla: **la persona esta enojada o feliz**
- Clasificación de especies de animales a partir de imágenes: **es un gato o un puma?**
- Detección de objetos en imágenes médicas:

Ejemplos: Classificacion

- Reconocimiento de emociones en el habla: **la persona esta enojada o feliz**
- Clasificación de especies de animales a partir de imágenes: **es un gato o un puma?**
- Detección de objetos en imágenes médicas: **es un tumor?**

Ejemplos: Classificacion

- Reconocimiento de emociones en el habla: **la persona esta enojada o feliz**
- Clasificación de especies de animales a partir de imágenes: **es un gato o un puma?**
- Detección de objetos en imágenes médicas: **es un tumor?**



Ejemplos: Regresion

- Reconocimiento de emociones en el habla:
- Evaluación de daños a partir de imágenes despues un terremoto:
- Detección de severidad de Alzheimer en la voz:

Ejemplos: Regresion

- Reconocimiento de emociones en el habla: **cual es la intesidad del enojo?**
- Evaluación de daños a partir de imágenes despues un terremoto:

- Detección de severidad de Alzheimer en la voz:

Ejemplos: Regresion

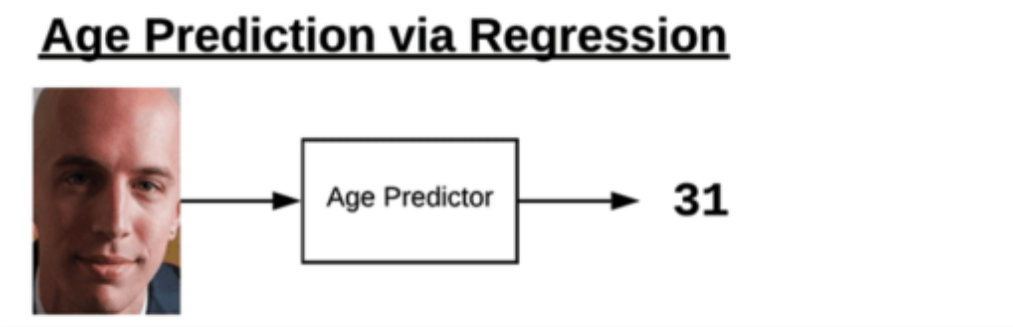
- Reconocimiento de emociones en el habla: **cual es la intesidad del enojo?**
- Evaluación de daños a partir de imágenes despues un terremoto: **la ambulancia puede utilizar el puente?**
- Detección de severidad de Alzheimer en la voz:

Ejemplos: Regresion

- Reconocimiento de emociones en el habla: **cual es la intesidad del enojo?**
- Evaluación de daños a partir de imágenes despues un terremoto: **la ambulancia puede utilizar el puente?**
- Detección de severidad de Alzheimer en la voz: **como avanzado es el estado?**

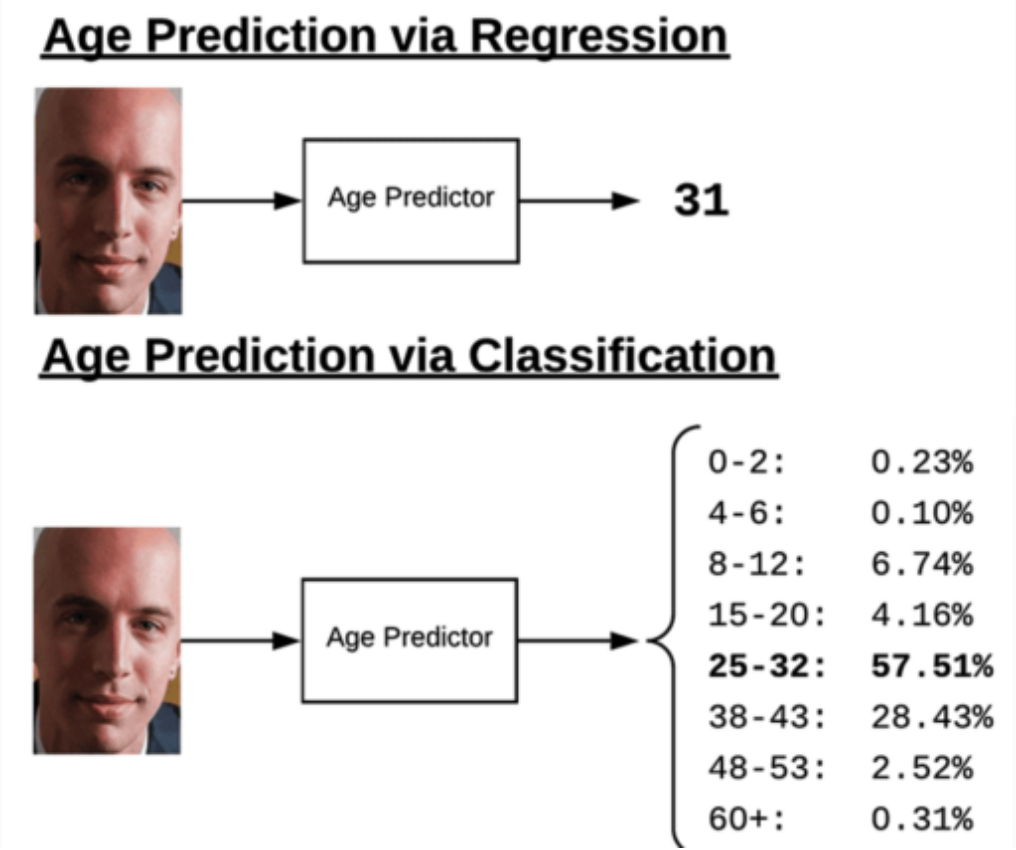
Ejemplos: Regresion

- Reconocimiento de emociones en el habla: **cual es la intesidad del enojo?**
- Evaluación de daños a partir de imágenes despues un terremoto: **la ambulancia puede utilizar el puente?**
- Detección de severidad de Alzheimer en la voz: **como avanzado es el estado?**



Ejemplos: Regresion

- Reconocimiento de emociones en el habla: **cual es la intesidad del enojo?**
- Evaluación de daños a partir de imágenes despues un terremoto: **la ambulancia puede utilizar el puente?**
- Detección de severidad de Alzheimer en la voz: **como avanzado es el estado?**



Ejemplos: Clustering

- Topic Mining en forums politicos:
- Clasificación de desinformación en redes sociales:
- Segmentación de clientes:

Ejemplos: Clustering

- Topic Mining en forums politicos: **de que los cuidadanos se preocupan?**
- Clasificación de desinformación en redes sociales:

- Segmentación de clientes:

Ejemplos: Clustering

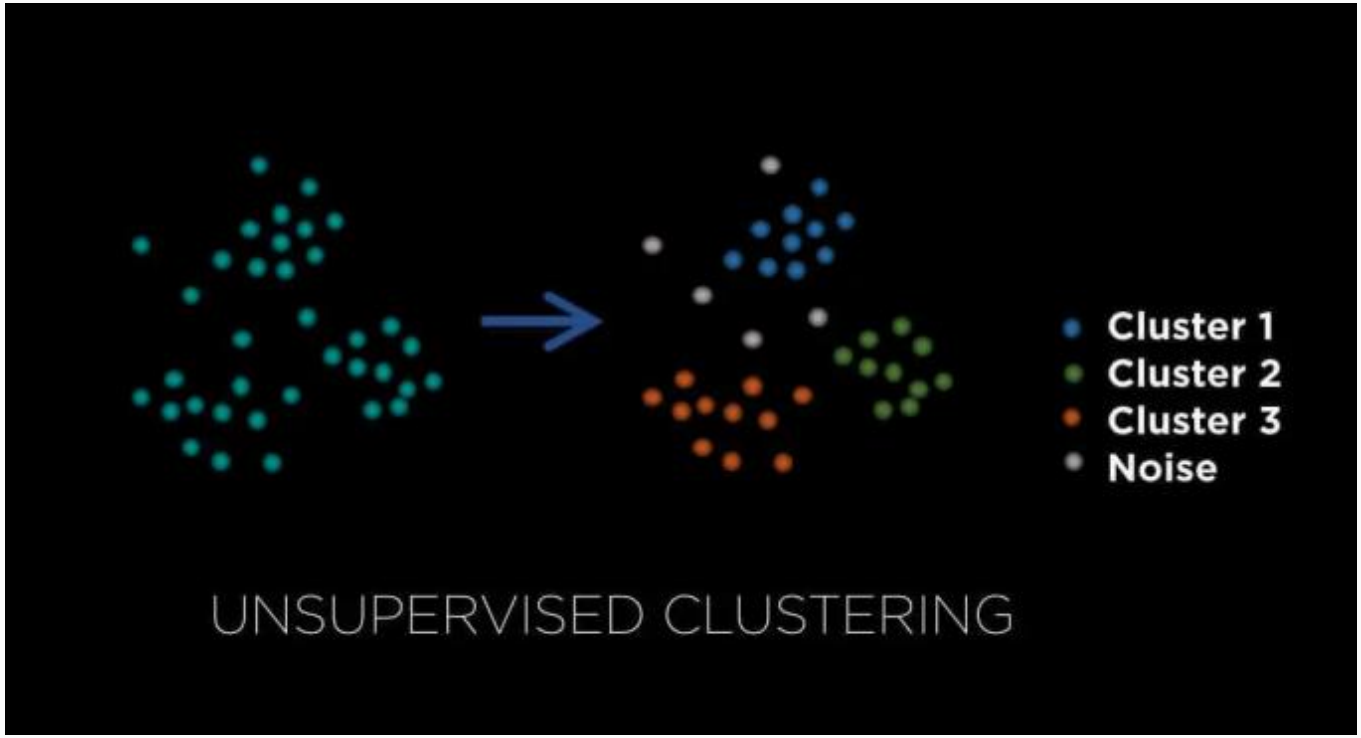
- Topic Mining en forums politicos: **de que los cuidadanos se preocupan?**
- Clasificación de desinformación en redes sociales: **agrupamos estas noticias que parecen raras**
- Segmentación de clientes:

Ejemplos: Clustering

- Topic Mining en forums politicos: **de que los cuidadanos se preocupan?**
- Clasificación de desinformación en redes sociales: **agrupamos estas noticias que parecen raras**
- Segmentación de clientes: **la gente que le gustan las chelas**

Ejemplos: Clustering

- Topic Mining en forums politicos: **de que los cuidadanos se preocupan?**
- Clasificación de desinformación en redes sociales: **agrupamos estas noticias que parecen raras**
- Segmentación de clientes: **la gente que le gustan las chelas**



Questions?